OPEN

Comparison of the Properties of Regression and Categorical Risk-Adjustment Models

Richard L. Fuller, MS; Richard F. Averill, MS;
John H. Muldoon, MHA; John S. Hughes, MD

Abstract: Clinical risk-adjustment, the ability to standardize the comparison of individuals with different health needs, is based upon 2 main alternative approaches: regression models and clinical categorical models. In this article, we examine the impact of the differences in the way these models are constructed on end user applications. **Key words:** *managed care, population health, reimbursement, risk-adjustment*

RISK-ADJUSTMENT MODELS are used in comparisons of hospitals, physician groups, and other health care organizations to account for differences in severity of illness and the complexity of care required. The selection and use of risk-adjustment models presents complicated technical and policy issues. Risk-adjustment is used to accomplish 5 basic policy objectives: (i) paying providers or health plans fairly when they assume financial risk; (ii) eliminating incentives to avoid complex and costly patients; (iii) standardizing comparisons of health care quality and patient outcomes; (iv) providing the basis for public policies that create the incentives to improve enrollee health and function; and (v) promoting transparent information exchange across all stakeholders, from policy makers to providers to patients. There are essentially 2 methods of constructing risk-adjustment models, regression and categorical. Both methods have been employed to achieve these 5 objectives with a wide range of applications of each method throughout the health care field.

We begin by outlining the differences in the approach used in the construction of regression and categorical risk-adjustment models. We then examine differences in how the 2 methods measure risk before discussing differences in applications of the 2 methods for management and policy, especially the issues of communication, transparency, and stability. We close by summarizing the key attributes of each type of risk-adjustment model relative to the 5 policy objectives.

CONSTRUCTION OF REGRESSION-BASED MODELS

Regression models are developed using statistical techniques to measure the response of

Author Affiliations: 3M Health Information Systems Silver Spring, Maryland (Mr Fuller, Mr Averill, and Mr Muldoon); Yale University School of Medicine, New Haven, Connecticut (Dr Hughes).

Correspondence: Richard L. Fuller, MS, c/o 12215 Plum Orchard Lane, Silver Spring, MD 20904 (rfuller@mmm.com).

157

outcomes of interest to predictor variables. In general, model development begins by employing a combination of literature search and expert panel knowledge to select factors believed to influence outcome variables of interest such as cost, complications, readmissions, or mortality.

Once selected, candidate predictor variables must be further specified so that they can be identified within available data sources. Thus, a decision to test the effect of chronic obstructive pulmonary disease (COPD) on patient cost first requires the creation of a variable to identify patients with COPD. The creation of the candidate predictor variables is a design decision affecting all subsequent results and may vary by developer. For example, developers must decide which *International Classification of Diseases, Tenth Revision,* codes constitute COPD, the sites of service from which reported data are utilized to identify COPD (eg, inpatient data only or both inpatient and outpatient data) and the time frame of the reported data used to identify COPD (eg, inpatient data only or inpatient data and data preceding the inpatient admission). After making these design decisions, the candidate predictor variables are tested within the framework of the statistical model.

The statistical framework is another design decision and can vary by both developer and the outcome being measured. For example, the Hierarchical Condition Category (HCC) capitation risk-adjustment model is built from a multiple linear regression that treats the relationship between observed enrollee costs and the predictor variables as primarily additive (Pope et al., 2004). Thus, with a few exceptions, the presence of a predictor variable adds a fixed amount to the prediction of enrollee cost. By contrast, the regression model developed for the Hospital Readmission Reduction Program (HRRP) is based upon the logistic (odds ratio) variant of the Hierarchical Generalized Linear Model, which is primarily multiplicative (Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation, 2014).

Thus, the presence of a predictor variable increases the predicted risk of readmission by a fixed percentage. The percentage increase is multiplicative to all other predictor variables, resulting in a compounding effect.

These design decisions are often made after testing the regression model to see which regression model offers superior fit (better prediction of the outcome). There is however a tendency for outcomes that are binary (yes/no) to be fitted with logistic regression models because of the possibility that standard regression models might predict an outcome with likelihood greater than 1 or less than 0 (asymptotically constrained). Because probabilities outside the range 0,1 are not meaningful, regression models for binary variables are first constrained within the probability range before optimizing data fit.

After specifying variables of interest and the statistical framework, the developer must specify other requirements, such as whether there will be any patient-level exclusions, service carve-outs, outlier trims or caps, etc. The developer can then calculate regression coefficients using statistical software. This entails using a development dataset complete with subsequent testing on a calibration, or validation dataset. Resulting coefficients and the overall fit of the model are subsequently reviewed for statistical significance.

In summary, regression models hypothesize variables that will predict an outcome, define how those variables are to be constructed, specify a statistical framework within which the predictor variables interact, compute fixed coefficient values for each predictor variable, and test whether there is sufficient association between each predictor variable and outcome for it to be retained within the regression model. The final model provides a formula from which the likelihood or magnitude of an outcome variable can be estimated given the presence of selected predictor variables (risk factors). Once developed, the regression formula can be used to predict the outcome in other databases, following the same approaches in terms of patient exclusions, service carve-outs, and outliers.

CONSTRUCTION OF CLINICAL CATEGORICAL MODELS

As with regression models, clinical categorical models utilize predictor variables to estimate the value of an outcome. The process used in the development of a clinical categorical model is an iterative process of formulating clinical hypotheses regarding the relationship between the outcome of interest and predictor variables and then testing the hypotheses with historical data. The historical data are used to confirm or refine the clinical hypotheses identified by clinicians. When there are discrepancies between clinical expectations and the data results from the historical data, the clinical expectations are refined to form the basis of the clinical categorical model. Typically, a clinical categorical model will be expressed as an exhaustive and mutually exclusive set of categories defined on the basis of the predictor variables with each category having a distinct predicted value of the outcome. The prediction of the outcome is based on the clinical category assigned and is typically computed as the average value of outcome for all patients assigned to the category. The most prominent example of a categorical model is diagnosis-related groups (DRGs).

Unlike regression models, categorical models require expert clinical opinion to hypothesize the likely effects (eg, large steps in cost or the presence of attenuating circumstances) of the predictor variables interaction with each other and the outcome of interest. How predictor variables interact is generally hypothesized to vary across health conditions; hence, a categorical model is hierarchically structured with layers of rules governing interactions among predictor variables.

It is important to highlight the number of potential interactions across predictor variables. The HCC regression model contains 111 distinct condition categories to predict enrollee cost within the Medicare Advantage risk-adjustment algorithm. Each variable carries with it a coefficient that can be added so as to predict enrollee cost. A categorical model developed using the same number of condition categories needs to account for 2^{111} possible interactions among these variables within its rules.

To place a limit upon the number of rules employed by the categorical model, the effects associated with variable interaction are constrained by the model's structure. For example, Medicare severity (MS)-DRGs are currently restricted to 749 payable MS-DRGs. Within these 749 categories, the impact of a particular predictor variable is neither assumed equal to its impact in another MS-DRG nor assumed to have a uniform impact when combined with other predictor variables within the MS-DRG. Instead, the combined impact from risk factors for cases within same DRG cells is hypothesized to have a similar impact upon the outcome variable. By allowing more cells, the categorical model allows a greater range of different interactions to be recognized. At the other end of the spectrum from using 749 DRGs to describe risk factor interaction for the Medicare Inpatient Prospective Payment System (IPPS), the Diagnosis Treatment Combinations used for payment in Holland have 29 000 distinct categories (Fleur, 2011). Although there is no magic number determining the number of cells that a categorical model should contain, the final model needs to be understandable to health professionals affected by the model. In addition, the final model needs to have sufficient volume distributed within cells so that the developers can assess the stability of the hypothesized relationships (without overfitting).

In summary, categorical models hypothesize variables that will predict outcomes, define how those variables are to be constructed, specify a clinical model with extensive rules that determine how the variables interact, and utilize data to test the clinical model at each stage as part of a process to confirm or revise the clinical rules. An in-depth description of the process of forming clinical categorical models can be found in the original article describing the formation of the DRGs (Thompson et al., 1979).

The end result is a clinical model that can be applied to many different population datasets from which weights or rates can be

separately calculated. For the clinical model to be broadly applicable to different populations, it is essential that the dataset used for testing include a full cross-section of the population, including less common but serious conditions. Users work from the same model, but are free to calculate weights or rates differently in terms of exclusions, carve-outs, outliers, etc. Unlike the coefficients in a regression model, the weights and rates used in a clinical model are independent of the underlying clinical model, allowing a uniform and stable clinical model to be maintained. For example, different payers, (eg, Medicare and Medicaid) can use different prices in payment applications of DRGs while calculating payment using the same DRGs.

In terms of similarities and differences, both models start by identifying variables of interest and operationalizing the definition of these variables from an available data source. The rest is quite different. The regression model specifies a statistical framework within which the variables interact, and computes a mathematical formula with coefficient values for the predictor variables, which are held constant across the model. The coefficients of the model are specific to a population dataset used to develop the regression formula. In contrast, the categorical model develops a clinical model with an extensive set of rules that specify the predictor variables and their varying interactions with each other and the outcome of interest. The final model is a clinical model that is separate from and can be applied to many different population datasets, from which weights and rates can then be calculated.

WHAT THE DIFFERENT APPROACHES TO CONSTRUCTION MEAN FOR MEASUREMENT OF RISK FACTORS

In practice, the decision to enforce the uniformity of effect on an outcome within a category (categorical) or to enforce uniformity of an effect by a variable across all interactions (regression) leads to significant differences in the measurement of risk factors. A general statement can be made that, in regression models, if the level of risk when 2 factors are present is greater than the sum of their individual risk then the calculated coefficients for individual risk factors will be overstated. Conversely, if the level of risk when the 2 factors are present is less than the sum of their individual risk, then the individual coefficients will be understated. A simplification of this difference is shown in Table 1.

In Table 1, we hypothesize 2 risk factors A and B that influence cost. In isolation, risk factor A is observed to increase cost by $100 and risk factor B by $300. When found together, cost is observed to increase by $1000 because of the interaction of the 2 variables. In the example we restrict the categorical model to 2 cells (high and low) and the regression to single terms only (no interactions). Furthermore, we assume that there are equal amounts of cases with risk factors A, B, and A/B. In this example, we assume the developers would be sufficiently proficient to structure the categorical model so as to keep the lower cost cases A and B (low) together, whereas the higher cost cases A/B (high) would form its own cell. The regression model will calculate coefficients for A ($300) and B ($500), with

Table 1. Difference Between Regression and Categorical Modeling of Risk

Risk Factor	Cost, $	Categorical Model	Cost, $	Regression Model	Cost, $
A	100	Low	200	A	300
B	300			B	500
A/B	1000	High	1000	A/B	800

the results summed to provide an estimate for cases with both A and B ($800).

This is, of course, a very simplified example. If it were this simple, a very accurate solution could easily be implemented. In the categorical model, a third cell could be added, and it would not be necessary to average the difference between just A and just B. In the regression model, an interactive term could be added to produce a distinct coefficient for cases that have both risk factors.

In actuality, outcomes of interest are a lot more complex and difficult to predict, and it is not a matter of simply adding additional interaction terms or categorical cells until prediction power is added. There are many reasons why developers restrict the number of terms or cells in a risk-adjustment model. First, by including a wider array of complex variables to capture interactions, one runs the risk of introducing more instability in the measured relationships between variables and predicted outcomes. This can be the result of "underpowered" cells (too few individuals with the matching array of conditions upon which to base a firm conclusion), colinearity across predictor variables (interaction effects between the predictors), clustering effects (with fewer observations exhibiting the characteristics of interest, there is a higher likelihood that they come from a single source, ie, hospital or region, thus embedding the performance of the source within the estimate of the effect), or simply random variation.

Second and related to the first reason, it is important to not overfit a risk-adjustment model to the development database. It is key that predictor variables have a clear clinical rationale and will hold up across datasets and over time (though periodically all models need to be reviewed and updated). To be successful, the model needs to capture the essential predictors, not omit any key predictors (omitted variable bias may give a false sense of precision to the coefficient or category estimates), and not include overlapping or redundant variables (which would allow for double-counting of risk variables and make the model very vulnerable to variations in coding and "code creep").

A third reason for restricting the number of predictor variables is that risk-adjustment models often need to avoid adjusting for unwanted and/or unnecessary variation. For example, increases in cost can be explained by the unnecessary use of expensive resources or complications from treatment. It is therefore reasonable to exclude those conditions that are predictive of cost but signal low-quality care from a risk-adjustment algorithm used for payment. Simply stated, a model that buries harm or poor performance within the algorithm, although it might improve explanatory power, provides no foundation for improvement.

Fourth, overly large unwieldy models with large numbers of interaction terms or multiple small volume cells confuse interpretation. Overall, those seeking to understand and redress cost and quality failings require objective, transparent, and understandable comparative information to move forward.

Developers therefore attempt to deliver parsimonious models to achieve the optimal tradeoff between maximizing predictive accuracy (consistent with policy objectives) and using the fewest predictors. The streamlining of models is not merely about the number of categories or variables but also the degree to which clinically meaningful distinctions can be drawn across them. This is particularly true of categorical-clinical models as they are inherently conceived as management tools, but is likewise important for the integrity of regression models.

WHAT THE DIFFERENT APPROACHES TO MODEL CONSTRUCTION MEANS FOR SYSTEM USE AND POLICY

In simpler situations where there are fewer interacting variables, regression and categorical models can yield similar results. That is, they can be used to predict similar amounts or rates of an outcome using the same predictor variables. This is not necessarily the case when interactions become more complex and the measurement of predictive accuracy also becomes more complex.

In addressing regression and categorical risk-adjustment models, it is important to understand risk and carefully evaluate predictive accuracy but it is also very important to evaluate a number of other attributes for system use and policy. Table 2 provides a comparison of design characteristics that should be considered and in the context of the 5 objectives of risk adjustment.

Table 2 begins by considering how differences in the method of model development affect the presentation of the model to users. A byproduct of creating a rule-based classification algorithm is that the rules used to classify patients and enrollees are available for review. Moreover, hierarchical classification results in discrete categories of patients (or enrollees) within understandable disease groupings. Together, these properties create a common clinical language for communication across stakeholders. This differs from regression models that permit a review of risk factor coefficients but only limited insight into *why* the interaction of patient risk factors manifests in a particular outcome. These properties are the cornerstone of transparency. Independent review of payment and performance measures is important for credibility, but transparency has an even greater role in letting clinicians know how their relative

Table 2. Comparison of Model Type Characteristics

Design Attribute	Clinical Categorical Model	Regression-Based Model
Development method	Clinical model developed by clinicians with formal classification rules governing assignment available for review	Statistical model developed with regression analysis
Structure of model	Clinically meaningful categories of enrollees subdivided into explicit severity of illness levels	Mathematical formula that computes a score
Communication value to providers	Creates a language understood by physicians	Numeric score that has minimal communication value
Calculation and replication of payment amounts	Arithmetic average which is easily calculated independent of developers	Requires regression analysis which can be difficult to perform independent of developers
Update process	Selective clinical areas can be refined without affecting entire clinical model	Requires respecification/re-estimation of statistical model
Response to changing practice patterns or technology	Clinical model stable but payment weights will change	Requires respecification/re-estimation of statistical model
Carve outs	Clinical model stable but payment weights will change	Requires respecification/re-estimation of statistical model
Model completeness	Clinical rules to define all potential interactions across predictor variables	Requires additional front-end logic to avoid redundancies (double counting and hierarchy) and formation of interaction variables (if used)

performance impacts both patient outcome and their own finances and areas in need of improvement.

Update, response to change, and accommodating policy variation (carve-outs) highlight how differences in the structure of the models affect their stability. Categorical clinical models are more robust in these circumstances as, after the initial extensive development, they are designed to tolerate changes within subsections of the model. As described previously, regression models are built with the requirement that risk factors have a common effect across all patient and disease subtypes, hence they require re-estimation and often respecification of the model in the face of change. The final component of Table 2 is model completeness. Regression models, such as HCCs, are not fully defined without the use of some categorical rules. Variables from the same family or those related by hierarchical severity require definition to avoid "double counting" and a subsequent unwanted mix of "paying for coding" and misestimation of their effect size. Moreover, the creation of the model usually identifies some interactions across variables that warrant an additional adjustment requiring definition.

The issues of transparency, communication, and stability can be understood using the HRRP risk-adjustment model as an example. The Centers for Medicare and Medicaid Services (CMS) segments performance measurement within the HRRP by admission type so that congestive heart failure (CHF) is modeled separately from pneumonia and acute myocardial infarction (AMI). Individual risk factor coefficients are permitted to vary across the models, thereby giving dissimilar signals, depending upon the admission type while holding the effect constant across all interactions within the model. For example, the presence of asthma or history of coronary artery bypass grafting is associated with reduced readmission risk for pneumonia admissions, whereas history of coronary artery bypass grafting is associated with increased readmission risk for AMI admissions and decreased risk for CHF admissions. The presence of asthma is associated with increased readmission risk for AMI and CHF admissions but not for all time periods. The variation in coefficient values across models indicates potential problems for a regression model that unifies readmission risk across admission types without first understanding or adjusting for the source of variation.

Stability of results over time, and thereby ability of users to identify important risk profiles, also requires interpretation. For example, dementia is associated with an increased readmission risk for AMI admissions in 1 year, decreased risk in the subsequent year, and little or no effect in a third. A similar pattern is produced by the effect of being male upon HF admissions (Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation, 2014).

A question posed by reviewing the HRRP model is how clinicians can use the risk-adjustment model to routinely identify those patients who are at the highest risk of readmission and are suffering from systematic failures. Another arguably more important question is whether or not these findings are clinically logical or more likely an artifact of data patterns that were somewhat random, particular to a database, or resulting from how the regression model was specified.

DISCUSSION

In the introduction to this article, we described 5 attributes of a risk-adjustment model upon which it should be compared and evaluated. The first 2 attributes, matching payment to patient risk and avoiding incentives to select less complex patients, are amenable to statistical measurement but require an understanding of their interrelatedness. Evaluating the accuracy of a model in matching payment to cost, or risk to likelihood of an outcome, has a tendency to rely upon an R^2 comparison of global fit. Although this may work with comparing overall historical patterns, it is essential to compare how well the model fits for subsections of a population—such as those with a particular chronic condition, of a particular location, with particular challenges or other identifiable risk factor. It should be

assumed that those paid to take on risk will look for ways to minimize their risk relative to their payment and will adopt strategies to do so, making this more complex analysis more important than the overall measure of global fit usually utilized to compare models.

Patient/enrollee subgroup risk is handled differently by the models. Regression models require more detailed and complex interaction terms to account for instances where patients/enrollees have either greater or lesser risk in the presence of an *additional* risk factor than would be predicted by simply adding the risk of the separate variables. Not doing so can lead to the transfer of allowance for risk *across* high- and low-risk case types for a given predictor variable. Categorical models are reliant upon the rules derived to classify patients/enrollees and therefore can benefit from very detailed clinical review. Classifying patients/enrollees within discrete cells will generally average the risk of some higher and lower risk patient types. Put differently, categorical models tend to transfer allowance for risk *within* high- or low-risk case types. This distinction between the 2 approaches is important when considering the impact upon providers. Two models may appear equivalent in their ability to discriminate between outcome predictions at an aggregate level, yet conceal a bias against particular providers because of the clustering of high-risk cases in particular providers or conversely by classifying distinct patient types of varying risk within single cells.

The attributes of communication, transparency, and stability are harder to quantify but easier to compare. Categorical models are designed to promote comparison, enable review of the rules that are generated to differentiate risk and to be modified in the face of expert commentary and insight. Categorical models separate the weighting of an outcome variable (such as cost) from the definition of patient complexity. The job of the model is to develop clinical rules to bring together similar patients/enrollees within a single consistent classification whereas the weight linking the model to risk adjustment (or payment) is computed and updated with changes in data and/or changes with regard to inclusions/exclusions, outlier policies, and other adjustment factors. As noted by the CMS the separation of the methodologies for developing the clinical model and the payment weights was a critical factor in the success and widespread adoption of the DRG system.

The separation of the clinical and payment weight methodologies allows stable clinical methodology to be maintained while the payment weights evolve in response to changing practice patterns.

Federal Register, May 4, 2001

But the separation of payment and classification introduced with the IPPS not only "set a reasonable price for a known product" (Schweiker, 1982) but also created a language to link the clinical and financial aspects of care, thereby improving communication between administrators and clinicians. Thus, a clinical categorical model can remain relatively stable, providing a consistent and powerful communication tool, whereas weights fluctuate to reflect changing practice patterns and new technology. The same model by being structured to define differences in patients/enrollees can be employed over varying geographic areas, over time and for a variety of outcomes without being dependent upon the initial data used in its development.

CONCLUSIONS

Regression and clinical categorical models represent very distinct approaches to risk adjustment. Users must carefully choose the model that best suites the intended application. Although clinical categorical models have many advantages in terms of communication, transparency, and stability, their initial development requires a significant effort and clinical input. Regression models usually require less initial development effort but are unstable in a changing environment and fail to provide the same degree of communication value and transparency.

REFERENCES

Fleur, H. (2011). *Incentives in the Diagnosis Treatment Combination payment system for specialist medical care: A study about behavioral responses of medical specialists and hospitals in the Netherlands*. Retrieved from http://www.promotiefleurhasaart.nl/pdf/IncentivesintheDiagnosisTreatmentCombinationSystemforspecialistmedicalcare_Hasaart.pdf

Pope, G. C., Kautter, J., Ellis, R. P., Ash, A. S., Ayanian, J. Z., Lezzoni, L. I., . . . Robst, J. (2004). Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financing Review*, *25*(4), 119–41. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4194896&tool=pmcentrez&rendertype=abstract

Schweiker, R. S. (1982). *Report to Congress: Hospital prospective payment for Medicare*. Retrieved from https://archive.org/details/reporttocongress00schw

Thompson, J. D., Averill, R. F., & Fetter, R. B. (1979). Planning, budgeting, and controlling—one look at the future: case-mix cost accounting. *Health Services Research*, *14*(2), 111–25. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1072108&tool=pmcentrez&rendertype=abstract

Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation. (2014). *2014 Measures updates and specifications report hospital-level risk-standardized readmission measures*. Retrieved from https://staging.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1219069855841